



# Performance Index for Tensor Polyadic Decompositions

Pierre Comon, Khalid Minaoui, Awatif Rouijel, Driss Aboutajdine

## ► To cite this version:

Pierre Comon, Khalid Minaoui, Awatif Rouijel, Driss Aboutajdine. Performance Index for Tensor Polyadic Decompositions. EUSIPCO 2013 - 21th European Signal Processing Conference, Sep 2013, Marrakech, Morocco. 5 p. hal-00842864

**HAL Id: hal-00842864**

**<https://hal.science/hal-00842864>**

Submitted on 9 Jul 2013

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# PERFORMANCE INDEX FOR TENSOR POLYADIC DECOMPOSITIONS

P. Comon\*

GIPSA-Lab, CNRS UMR5216,  
Grenoble Campus, BP.46,  
F-38402 St Martin d'Heres cedex, France

K. Minaoui, A. Rouijel, D. Aboutajdine<sup>†</sup>

University Mohamed V-Agdal,  
LRIT Associated Unit to the CNRST-URAC N29,  
avenue des nations unies, Agdal, BP.554,  
Rabat-Chellah, Morocco

## ABSTRACT

It is proposed to isolate the computation of the scaling matrix in CP tensor decompositions. This has two implications. First, the conditioning of the problem shows up explicitly, and could be controlled via a constraint on the so-called coherences. Second, a performance measure concerning only the factor matrices can be exactly calculated, and does not present the optimistic bias of the minimal error generally utilized in the literature. In fact, for tensors of order  $d$ , it suffices to solve a degree-2 polynomial system in  $d$  variables. We subsequently give an explicit solution when  $d = 3$ .

**Index Terms**—Blind ; Source separation ; Data mining ; Identification ; Array ; Tensor ; CP ; Canonical polyadic ; Candecomp ; Parafac ; Performance ; Coherence

## 1. INTRODUCTION

Tensor decompositions are now seen as promising tools in data mining [1, 2, 3] and signal processing [4, 5, 6, 7], to cite a few. In [4] for instance, a deterministic approach has been proposed, which permits not only to work with short data lengths, to localize more sources than sensors, but also to extract source copies for free.

In tensor-based approaches, it is well known that factor matrices are identified up to column scaling. This indeterminacy is complicated to take into account because of a constraint: the product of all scaling matrices should be equal to identity. For this reason, only approximate performance indices have been used so far, by simply ignoring the latter constraint. In this paper, we concentrate on calculating an exact performance index, and we subsequently show that the usual approximate index can be significantly optimistic. This is illustrated by computer experiments involving three decomposition algorithms, including two new ones, which are interesting in and of themselves.

\*Work funded by the European Research Council under the European Community's Seventh Framework Programme FP7/2007–2013 Grant Agreement no. 320594.

<sup>†</sup>The 2nd and 3rd authors contributed while visiting Gipsa-Lab.

## 2. NOTATION

In the following, vectors will be denoted by bold lowercases, *e.g.*  $\mathbf{a}$ , whereas matrices or higher-order arrays will be denoted by bold uppercases, *e.g.*  $\mathbf{A}$ . Moreover,  $\mathbf{a}_r$  will denote the  $r$ th column of matrix  $\mathbf{A}$ ,  $\text{diag } \boldsymbol{\lambda}$  the diagonal matrix whose diagonal entries are  $\lambda_i$ ,  $\mathbf{1}$  will represent a vector containing ones.

We are interested in decomposing a 3rd order tensor  $\mathbf{T}$  as:

$$\mathbf{T} = \sum_{r=1}^R \lambda_r \mathbf{E}(r) \quad (1)$$

where  $\mathbf{E}(r)$  are *decomposable* tensors, that is,  $\mathbf{E}(r) = \mathbf{a}_r \otimes \mathbf{b}_r \otimes \mathbf{s}_r$ , where  $\otimes$  denotes the tensor (outer) product, and  $R$  is the *tensor rank*. Vectors  $\mathbf{a}_r$  (resp.  $\mathbf{b}_r$  and  $\mathbf{s}_r$ ) live in a linear space of dimension  $I$  (resp. dimension  $J$  and  $K$ ). Once bases of these three spaces are fixed, tensor decomposition (1) is equivalent to the decomposition of the  $I \times J \times K$  array of coordinates:

$$T_{ijk} = \sum_{r=1}^R \lambda_r A_{ir} B_{jr} S_{kr} \quad (2)$$

where  $A_{ir}$  (resp.  $B_{jr}$  or  $S_{kr}$ ) denote the entries of vector  $\mathbf{a}_r$  (resp.  $\mathbf{b}_r$  or  $\mathbf{s}_r$ ). Equation (1) is often referred to as the Canonical Polyadic decomposition (CP)<sup>1</sup> of  $\mathbf{T}$ . Even when the CP is unique, the explicit writing of decomposable tensors in given bases as in (2) is subject to scale indeterminacies. By normalizing the  $R$  columns of matrices  $\mathbf{A}$ ,  $\mathbf{B}$  and  $\mathbf{S}$ , *i.e.* vectors  $\mathbf{a}_r$ ,  $\mathbf{b}_r$  and  $\mathbf{s}_r$ , and by pulling real positive factors  $\lambda_r$  outside the product, these indeterminacies are clearly reduced to unit modulus but are not completely fixed, hence the difficulty in estimating the identification error of factor matrices  $\mathbf{A}$ ,  $\mathbf{B}$  and  $\mathbf{S}$ . The purpose of Section 4 is precisely to fix these  $3R$  complex phases (reducing to signs in the real case).

## 3. EXISTENCE AND UNIQUENESS

The goal is to identify all parameters in the right hand side of (2), given the whole array  $\mathbf{T}$ . According to existing results [8,

<sup>1</sup>also sometimes called Candecomp/Parafac in Psychometry.

9], uniqueness<sup>2</sup> of the exact CP decomposition (2) is generally ensured when  $I + J + K \geq 2R + 2$  (see infra for more details); and this condition is claimed to be only sufficient for  $R > 3$ . However, observations are actually corrupted by noise, so that (2) does not hold exactly.

### 3.1. Low rank approximation

The natural idea is then to fit model (2) by minimizing the error

$$\Upsilon(\mathbf{A}, \mathbf{B}, \mathbf{S}; \mathbf{\Lambda}) = \|\mathbf{T} - (\mathbf{A}, \mathbf{B}, \mathbf{S}) \cdot \mathbf{\Lambda}\|^2 \quad (3)$$

where  $\mathbf{\Lambda}$  denotes the  $R \times R \times R$  diagonal array whose diagonal entries are  $\lambda_r$ , and where  $(\mathbf{A}, \mathbf{B}, \mathbf{S}) \cdot \mathbf{\Lambda}$  is an Ersatz for the three-way tensor of coordinates  $\sum_r \lambda_r A_{ir} B_{jr} S_{kr}$ .

Up to now, the choice of the norm has not been specified, but we shall subsequently use the Frobenius norm, for either matrices or tensors:  $\|\mathbf{T}\|^2 = \sum_{ijk} |T_{ijk}|^2$ , and the  $L^2$  norm for vectors.

Minimizing error (3) means finding the best rank- $R$  approximate of  $\mathbf{T}$  and its CP decomposition. However, as already pointed out in [10, 11] and references therein, the infimum of  $\Upsilon$  may not always be reached, even if it often is. But this is out of the scope of the present paper.

### 3.2. Coherence

In compressed sensing [12, 13] the *coherence* of a set of unit norm vectors is defined as the maximal value of the modulus of cross scalar products:

$$\mu_A = \sup_{p \neq q} |\mathbf{a}_p^H \mathbf{a}_q| \quad (4)$$

We define this way coherences  $\mu_A$ ,  $\mu_B$  and  $\mu_S$  associated with matrices  $\mathbf{A}$ ,  $\mathbf{B}$  and  $\mathbf{S}$ , respectively, being understood that  $\mathbf{a}_r$ ,  $\mathbf{b}_r$  and  $\mathbf{s}_r$  denote their columns.

Coherences play a role in the conditioning of the problem. To see this, suppose matrices  $\mathbf{A}$ ,  $\mathbf{B}$  and  $\mathbf{S}$  are given. Then the optimal value  $\mathbf{\Lambda}^o$  minimizing error  $\Upsilon$  satisfies the following linear system:

$$\mathbf{G} \mathbf{\Lambda}^o = \mathbf{f}, \quad (5)$$

where  $\mathbf{G}$  denotes the  $R \times R$  Gram matrix defined by:

$$G_{pq} = (\mathbf{a}_p \otimes \mathbf{b}_p \otimes \mathbf{s}_p)^H (\mathbf{a}_q \otimes \mathbf{b}_q \otimes \mathbf{s}_q),$$

$\mathbf{\Lambda}$  is the  $R$ -dimensional vector with entries  $\lambda_r$ ,  $\otimes$  is the Kronecker product and vector  $\mathbf{f}$  in the RHS is defined by the contraction  $f_r = \sum_{ijk} T_{ijk} A_{ir}^* B_{jr}^* S_{kr}^*$ ,  $1 \leq r \leq R$ . This can be seen by expanding the Frobenius norm in (3), which is a quadratic form in the entries of  $\mathbf{\Lambda}$ , and by canceling the gradient w.r.t.  $\mathbf{\Lambda}$ .

In view of (5), and since diagonal entries of  $\mathbf{G}$  are equal to 1, it is clear that imposing coherences to have a modulus strictly smaller than 1 will permit to impose an acceptable

<sup>2</sup>Uniqueness is always understood up to matrix scale factors.

conditioning. Also note the striking fact that only the *product* between coherences appears, and not coherences individually. It turns out that this statement has deeper implications, especially in existence and uniqueness of the solution to Problem (3), as briefly elaborated below.

### 3.3. Existence

It has been shown in [11, 14] that if

$$\mu_A \mu_B \mu_S < \frac{1}{R-1} \quad (6)$$

then the infimum of (3) is reached. The reason for this is that error (3) becomes coercive as soon as (6) is satisfied. And since it is continuous, it must reach its minimum. We see that this condition already gives a quantitative bound to the conditioning of (5).

### 3.4. Uniqueness

Now to give a sufficient condition to uniqueness, we can rely on Kruskal's theorem previously quoted, which can be expressed as follows:  $\mathbf{T} = \sum_{r=1}^R \mathbf{E}(r)$  where  $\mathbf{E}(r)$  are rank-one tensors, admits a unique solution if

$$\text{krank}\{\mathbf{A}\} + \text{krank}\{\mathbf{B}\} + \text{krank}\{\mathbf{S}\} \geq 2R + 2 \quad (7)$$

where  $\text{krank}\{\cdot\}$  denotes Kruskal's rank<sup>3</sup>.

Following the lines of [12, 11, 14], one can observe that  $\text{krank}\{\mathbf{A}\} \geq \mu_A^{-1}$ , as long as  $\text{krank}\{\mathbf{A}\}$  is strictly smaller than the column rank of  $\mathbf{A}$ . Plugging this inequality in (7) leads to the sufficient uniqueness condition:

$$\mu_A^{-1} + \mu_B^{-1} + \mu_S^{-1} \geq 2R + 2 \quad (8)$$

## 4. PERFORMANCE MEASURE

As pointed out in section 2, there remains an indeterminacy in the CP decomposition, characterized by  $2R$  complex numbers of unit modulus. More precisely, let  $\mathbf{a}$ ,  $\mathbf{b}$  and  $\mathbf{s}$  denote the  $r$ th column of matrices  $\mathbf{A}$ ,  $\mathbf{B}$  and  $\mathbf{S}$ , respectively,  $r$  being any fixed value between 1 and  $R$ . Also denote  $\hat{\mathbf{a}}$ ,  $\hat{\mathbf{b}}$  and  $\hat{\mathbf{s}}$  one column of the estimated matrices entering in the CP decomposition. It is desired to define the minimal distance:

$$\delta(\mathbf{x}; \hat{\mathbf{x}}) = \min_{\varphi, \psi, \chi} \{ \|\mathbf{a} - e^{j\varphi} \hat{\mathbf{a}}\|^2 + \|\mathbf{b} - e^{j\psi} \hat{\mathbf{b}}\|^2 + \|\mathbf{s} - e^{j\chi} \hat{\mathbf{s}}\|^2 \} \quad (9)$$

under the constraint that  $\exp(j(\varphi + \psi + \chi)) = 1$ . For convenience,  $\mathbf{x}$  again denotes the vector  $[\mathbf{a}^T, \mathbf{b}^T, \mathbf{s}^T]^T$ . It turns out that this distance can be exactly computed. In fact, because of the constraint, (9) involves two angles, say  $\varphi$  and  $\psi$ . An

<sup>3</sup> $k_A = \text{krank} \mathbf{A}$  if any  $k_A$  columns of matrix  $\mathbf{A}$  are linearly independent. This is in contrast to the usual matrix rank, where *any* is replaced by *at least* in the definition.

elimination is possible in the two equations defining stationary points, and leads to the roots of a polynomial of degree 6 in a single variable, say  $\varphi$ . By plugging back the admissible values of  $\varphi$  in the system of two equations, corresponding values of  $\psi$  are obtained. The global minimum of (9) can hence be efficiently calculated (see appendix for details).

We end up with the performance criterion below:

$$\mathcal{E}(\mathbf{T}; \mathbf{A}, \mathbf{B}, \mathbf{S}, \mathbf{\Lambda}) = \min_{\pi \in \Pi} \sum_{r=1}^R \delta(\mathbf{x}_r; \hat{\mathbf{x}}_{\pi(r)}) \quad (10)$$

where  $\Pi$  is the group of permutations of  $\{1, 2, \dots, R\}$ . This criterion allows to properly fix the permutation-scale ambiguity. In case the permutation acts in too large dimension, greedy versions are possible to limit the exhaustive search in the permutation group.

Surprisingly, such a performance criterion using the solution of (9) has not yet been proposed in the literature.

## 5. COMPUTER RESULTS

### 5.1. Optimization

In order to illustrate the interest of our performance measure, we shall report the performances of three simple algorithms computing the approximate CP decomposition: Algorithm 0 is a gradient descent, where variables  $\mathbf{A}$ ,  $\mathbf{B}$  and  $\mathbf{S}$  are free, and  $\mathbf{\Lambda} = \mathbf{I}$ ; Algorithm 1 is a projected gradient with an approximate value of  $\mathbf{\Lambda}$ ; Algorithm 2 is a projected gradient with the optimal value of  $\mathbf{\Lambda}$  given by (5). Note that the focus is on the performance measure, not on algorithms. For the sake of clarity, Algorithms 1 and 2 are detailed below.

ALGORITHM 1.

1. Initialize  $(\mathbf{A}(0), \mathbf{B}(0), \mathbf{S}(0))$  to full-rank matrices with unit-norm columns, and set  $\boldsymbol{\lambda}(0) = \mathbf{1}$ .
2. For  $k \geq 1$  and subject to a stopping criterion, do
  - (a) Compute the descent direction as the gradient w.r.t.  $\mathbf{X}$ :  $\mathbf{D}(k) = -\nabla \Upsilon(\mathbf{X}(k-1); \boldsymbol{\lambda}(k-1))$
  - (b) Compute a stepsize  $\ell(k)$
  - (c) Update  $\mathbf{X}(k) = \mathbf{X}(k-1) + \ell(k) \mathbf{D}(k)$
  - (d) Extract the 3 blocks of  $\mathbf{X}(k)$ :  $\mathbf{A}(k)$ ,  $\mathbf{B}(k)$  and  $\mathbf{S}(k)$ , and store the norm of their columns into  $\boldsymbol{\lambda}_A$ ,  $\boldsymbol{\lambda}_B$ ,  $\boldsymbol{\lambda}_S$
  - (e) Normalize them to unit-norm columns as  $\mathbf{A}(k) := \mathbf{A}(k) \text{diag}(\boldsymbol{\lambda}_A)^{-1}$ ,  $\mathbf{B}(k) := \mathbf{B}(k) \text{diag}(\boldsymbol{\lambda}_B)^{-1}$  and  $\mathbf{S}(k) := \mathbf{S}(k) \text{diag}(\boldsymbol{\lambda}_S)^{-1}$
  - (f) update  $\boldsymbol{\lambda}(k) = \boldsymbol{\lambda}(k-1) * \boldsymbol{\lambda}_A * \boldsymbol{\lambda}_B * \boldsymbol{\lambda}_S$

where  $*$  denotes the entry-wise product.

ALGORITHM 2.

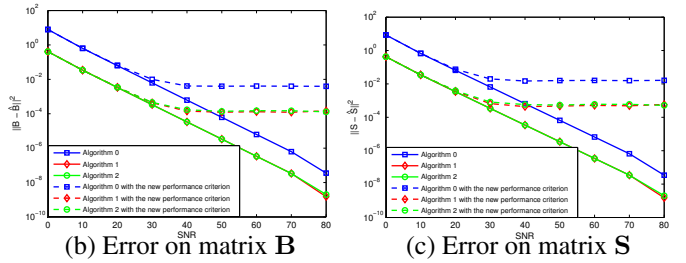
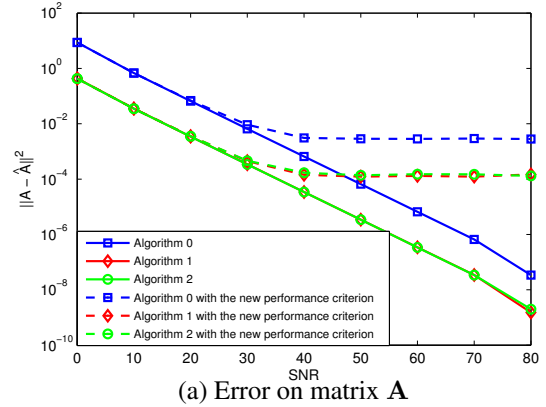
1. Initialize  $(\mathbf{A}(0), \mathbf{B}(0), \mathbf{S}(0))$  to full-rank matrices with unit-norm columns.
2. Compute  $\mathbf{G}(0)$  and  $\mathbf{f}(0)$ , and solve  $\mathbf{G}(0) \boldsymbol{\lambda}(0) = \mathbf{f}(0)$  for  $\boldsymbol{\lambda}$ , as defined in Section 3.2.
3. For  $k \geq 1$  and subject to a stopping criterion, do
  - (a) Compute the descent direction as the gradient w.r.t.  $\mathbf{X}$ :  $\mathbf{D}(k) = -\nabla \Upsilon(\mathbf{X}(k-1); \boldsymbol{\lambda}(k-1))$

- (b) Compute a stepsize  $\ell(k)$
- (c) Update  $\mathbf{X}(k) = \mathbf{X}(k-1) + \ell(k) \mathbf{D}(k)$
- (d) Extract the 3 blocks of  $\mathbf{X}(k)$ :  $\mathbf{A}(k)$ ,  $\mathbf{B}(k)$  and  $\mathbf{S}(k)$
- (e) Normalize the columns of  $\mathbf{A}(k)$ ,  $\mathbf{B}(k)$  and  $\mathbf{S}(k)$
- (f) Compute  $\mathbf{G}(k)$  and  $\mathbf{f}(k)$ , and solve  $\mathbf{G}(k) \boldsymbol{\lambda}(k) = \mathbf{f}(k)$  for  $\boldsymbol{\lambda}$ , according to (5).

The gradient expressions<sup>4</sup> necessary to determine the descent direction  $\mathbf{D}(k)$  are of the form:

$$\frac{\partial \Upsilon}{\partial \mathbf{A}} = 2 \mathbf{A} \mathbf{M} - 2 \mathbf{N}$$

where  $M_{pq} \stackrel{\text{def}}{=} \sum_{jk} \lambda_p B_{jp} S_{kp} S_{kq}^* B_{jq}^* \lambda_q^*$  and  $N_{ip} \stackrel{\text{def}}{=} \sum_{jk} T_{ijk} B_{jp}^* S_{kp}^* \lambda_p^*$ . Expressions for gradients w.r.t.  $\mathbf{B}$  or  $\mathbf{S}$  are similar.



**Fig. 1.** Matrix estimation errors, with a random tensor of size  $10 \times 10 \times 10$  and rank 5: (a) matrix  $\mathbf{A}$ , (b) matrix  $\mathbf{B}$ , (c) matrix  $\mathbf{S}$ . Note the 3 matrix factors typically contribute equivalently in  $\mathcal{E}$ .

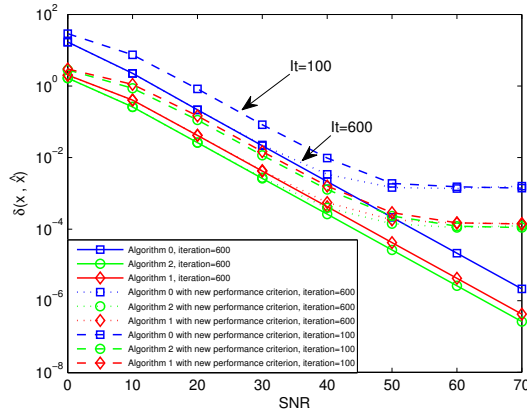
### 5.2. Performances

In this section, we analyze matrix estimation errors obtained with the performance criterion proposed in (9). Two scenarios are analyzed with random tensors of rank 5 corrupted by an additive Gaussian noise: one with dimensions  $4 \times 4 \times 4$  and another with dimensions  $10 \times 10 \times 10$ . Factor matrices are initialized with random values with 5 columns. The results

<sup>4</sup>Matrix gradients are written with the conventions described in [15, 16].

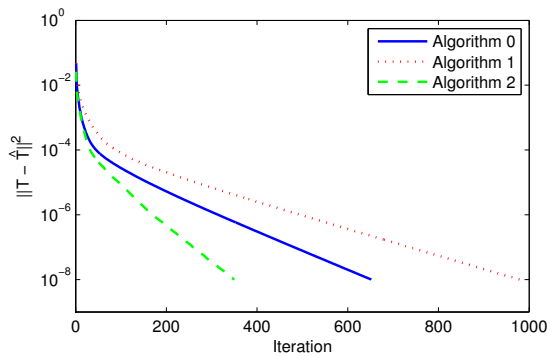
were obtained from 50 Monte Carlo runs. At each run and for every SNR value, a new noise realization is drawn.

Figures 1 and 2 report matrix estimation errors involved in error (9) as a function of SNR. First, it can be seen that the performance results using Algorithm 0 is poor in comparison with results obtained with Algorithms 1 and 2 (curves with stars and curves circles). This supports the idea that our algorithms isolating the scale matrix are attractive. Second, we check that when we do not consider the phase constraints when calculating the performance measure, the results are significantly more optimistic, especially at high SNR (compare curves with the same color/symbol); this supports the interest in using our performance index defined in (9).



**Fig. 2.** Sum  $\mathcal{E}$  of matrix estimation errors, with a random tensor of size  $4 \times 4 \times 4$  and rank 5. Note the asymptote depending on the maximum number of iterations executed.

Third, to see a significant difference between Algorithms 1 and 2, it is necessary to look at the convergence speed, since the final error is about the same (cf. Figures 1 and 2). In all our experiments, Algorithm 2 converged faster in terms of number of iterations; this is illustrated in Figure 3.



**Fig. 3.** Typical example of reconstruction error (3) as a function of the number of iterations, for a tensor of size  $4 \times 4 \times 4$  and rank 5.

## 6. CONCLUDING REMARKS

We have shown in Section 3.2 that, in CP tensor decompositions, the scale matrix  $\mathbf{A}$  takes as optimal value a Gram matrix controlling the conditioning of the problem. This shows that bounding coherences would allow to ensure a minimal conditioning. Next, two descent algorithms have been described and tested, which involve a separate explicit calculation of the scale matrix  $\mathbf{A}$ . Third, this numerical approach allows to compute a performance index, more realistic than performance measures used in the literature which are optimistic by construction. Future works include the development of more efficient numerical algorithms based on the same idea, with possibly a coherence constraint.

## 7. APPENDIX

In this appendix, we explain in more details how one can obtain performance index  $\delta$ , and in particular how phases  $(\varphi, \psi, \chi)$  are calculated. Setting  $\chi = -\varphi - \psi [2\pi]$ , equation (9) can be rewritten as:

$$\begin{aligned} \delta = & \|\mathbf{a}\|^2 + \|\hat{\mathbf{a}}\|^2 + \|\mathbf{b}\|^2 + \|\hat{\mathbf{b}}\|^2 + \|\mathbf{s}\|^2 + \|\hat{\mathbf{s}}\|^2 \\ & - 2\rho_a \cos(\varphi - \alpha) - 2\rho_b \cos(\psi - \beta) \\ & - 2\rho_s \cos(\varphi + \psi + \gamma) \end{aligned}$$

where  $\mathbf{a}^H \hat{\mathbf{a}} \stackrel{\text{def}}{=} \rho_a e^{j\alpha}$ ,  $\mathbf{b}^H \hat{\mathbf{b}} \stackrel{\text{def}}{=} \rho_b e^{j\beta}$  and  $\mathbf{s}^H \hat{\mathbf{s}} \stackrel{\text{def}}{=} \rho_s e^{j\gamma}$ . Stationary points are given by the solutions of the trigonometric system in *e.g.* variables  $x = \varphi - \alpha$  and  $y = \psi - \beta$  as unknowns:

$$\begin{aligned} \rho_a \sin x + \rho_s \sin(\varphi + \psi + \gamma) &= 0 \\ \rho_b \sin y + \rho_s \sin(\varphi + \psi + \gamma) &= 0 \end{aligned}$$

The first simplification is achieved by noting that

$$\rho_s \sin(\varphi + \psi + \gamma) = -\rho_a \sin x = -\rho_b \sin y.$$

implies  $\sin y = \sin x \frac{\rho_a}{\rho_b}$ . Now, using trigonometric identities, we can rewrite the first equation of the trigonometric system

$$\rho_s \sin(\varphi + \psi + \gamma) = \rho_s \sin(x + y + \alpha + \beta + \gamma) = -\rho_a \sin x,$$

$$\begin{aligned} \text{as: } \rho_a \sin x = & -\rho_s [\sin x \cos y \cos(\alpha + \beta + \gamma) \\ & + \sin y \cos x \cos(\alpha + \beta + \gamma) \\ & + \cos x \cos y \sin(\alpha + \beta + \gamma) \\ & - \sin x \sin y \sin(\alpha + \beta + \gamma)]. \end{aligned}$$

Letting  $\cos y = \sqrt{1 - \sin^2 y}$  and  $\sin y = \frac{\rho_a}{\rho_b} \sin x$ , we obtain

$$\begin{aligned} \rho_a \sin x = & -\frac{\rho_s \rho_a}{\rho_b} \sin x \cos x \cos(\alpha + \beta + \gamma) \\ & + \frac{\rho_s \rho_a}{\rho_b} \sin^2 x \sin(\alpha + \beta + \gamma) \\ & + [\rho_s \sin x \sin(\alpha + \beta + \gamma) \\ & - \rho_s \cos x \sin(\alpha + \beta + \gamma)] \sqrt{1 - \frac{\rho_a}{\rho_b} \sin^2 x}. \end{aligned}$$

The goal of the next step is to eliminate the square root and to rewrite the equation in term of variables  $\sin x$  or  $\cos x$ . So, let us squaring both side of this equation and using trigonometric identities such as  $\cos^2 x = \frac{1+\cos(2x)}{2}$ ,  $\sin^2 x = \frac{1-\cos(2x)}{2}$ ,  $\cos x \sin x = \frac{\sin(2x)}{2}$  and  $\cos^2(2x) + \sin^2(2x) = 1$ . Thus, after simplification we obtain

$$\begin{aligned} & \frac{\rho_b^2}{2} + \frac{1}{2} \left( \frac{\rho_s \rho_a}{\rho_b} \right)^2 - \frac{\rho_s^2}{2} + \left[ \frac{\rho_b^2}{2} + \frac{1}{2} \left( \frac{\rho_s \rho_a}{\rho_b} \right)^2 \right. \\ & + \frac{\rho_s^2}{2} \cos^2(\alpha + \beta + \gamma) - \frac{\rho_s^2}{2} \sin^2(\alpha + \beta + \gamma) \left. \right] \cos^2(2x) \\ & + \left[ 2\rho_s \cos(\alpha + \beta + \gamma) \sin(\alpha + \beta + \gamma) \right] \sqrt{1 - \cos^2(2x)} \\ & + \left[ 2 \left( \frac{\rho_s \rho_a^2}{\rho_b} \right) \cos(\alpha + \beta + \gamma) \sqrt{1 - \cos^2(2x)} \right. \\ & \left. - \left( \frac{\rho_s \rho_a^2}{\rho_b} \right) \sin(\alpha + \beta + \gamma) (1 - \cos(2x)) \right] \sqrt{\frac{1 - \cos(2x)}{2}} \\ & = 0. \end{aligned}$$

In the same way as above, we squared twice both sides of the resulting equation to eliminate the squares roots. Finally we get an equation of degree six of the form

$$c_0 + c_1 \cos(2x) + c_2 \cos^2(2x) + c_3 \cos^3(2x) + c_4 \cos^4(2x) + c_5 \cos^5(2x) + c_6 \cos^6(2x) = 0,$$

$$\text{with: } \begin{cases} c_0 = c_1'^2 - c_5'^2 \\ c_1 = 2c_1'c_4' - 2c_5'c_7' \\ c_2 = 2c_1'c_3' + c_4'^2 - 2c_5'c_6' - c_7'^2 + c_5'^2 \\ c_3 = 2c_1'c_2' + 2c_3'c_4' - 2c_6'c_7' + 2c_5'c_7' \\ c_4 = c_3'^2 + 2c_2'c_4' - c_6'^2 + 2c_5'c_6' + c_7'^2 \\ c_5 = 2c_2'c_3' + 2c_6'c_7' \\ c_6 = c_2'^2 + c_6'^2 \\ c_1' = c_1''^2 + c_3''^2 - \frac{1}{2}c_4''^2 - \frac{1}{2}c_5''^2 \\ c_2' = -\frac{1}{2}c_4''^2 + \frac{1}{2}c_5''^2 \\ c_3' = c_2''^2 - c_3''^2 + \frac{1}{2}c_4''^2 - \frac{3}{2}c_5''^2 \\ c_4' = 2c_1''c_2' + \frac{1}{2}c_4''^2 + \frac{4}{2}c_5''^2 \\ c_5' = -2c_1''c_3' + c_4''c_5'' \\ c_6' = c_4''c_5'' \\ c_7' = -2c_2''c_3' - 2c_4''c_5'' \end{cases}$$

$$\text{and: } \begin{cases} c_1'' = \frac{1}{2}\rho_a^2 + \frac{1}{2} \left( \frac{\rho_s \rho_a}{\rho_b} \right)^2 - \frac{1}{2}\rho_s^2 \\ c_2'' = -\frac{1}{2}\rho_a^2 - \frac{1}{2} \left( \frac{\rho_s \rho_a}{\rho_b} \right)^2 + \rho_s^2 \cos^2(\alpha + \beta + \gamma) - \frac{1}{2} \\ c_3'' = -2\rho_s^2 \cos(\alpha + \beta + \gamma) \sin(\alpha + \beta + \gamma) \\ c_4'' = -2 \frac{\rho_s \rho_a^2}{\rho_b} \cos(\alpha + \beta + \gamma) \\ c_5'' = \frac{\rho_s \rho_a^2}{\rho_b} \sin(\alpha + \beta + \gamma) \end{cases}$$

Solving the sixth degree equation yields  $x$ . Replacing  $x$  in  $\sin y = \frac{\rho_a}{\rho_b} \sin x$  yields  $y$ .

#### Acknowledgment

As pointed out by a reviewer, our criterion may be related to the squared angular error whose bounds have been derived in [17].

## 8. REFERENCES

- [1] A. Smilde, R. Bro, and P. Geladi, *Multi-Way Analysis*, Wiley, Chichester UK, 2004.
- [2] M. Mørup, "Applications of tensor (multiway array) factorizations and decompositions in data mining," *Data Mining and Knowledge Discovery*, vol.1, no.1, pp. 24–40, 2011.
- [3] D. M. Dunlavy, T. G. Kolda, and E. Acar, "Temporal link prediction using matrix and tensor factorizations," *ACM Transactions on Knowledge Discovery from Data*, vol.5, no.2, Feb. 2011.
- [4] N. D. Sidiropoulos, R. Bro, and G. B. Giannakis, "Parallel factor analysis in sensor array processing," *IEEE Trans. Sig. Proc.*, vol.48, no.8, pp. 2377–2388, Aug. 2000.
- [5] J. M. Landsberg, *Tensors: Geometry and Applications*, vol.128 of *Graduate Studies in Mathematics*, AMS publ., 2012.
- [6] E. Acar, T. G. Kolda, D. M. Dunlavy, and M. Mørup, "Scalable tensor factorizations for incomplete data," *Chemometrics Intel. Lab. Syst.*, 2011.
- [7] P. Comon and C. Jutten, Eds., *Handbook of Blind Source Separation, Independent Component Analysis and Applications*, Academic Press, 2010. hal-00460653.
- [8] J. M. F. Ten Berge and N. Sidiropoulos, "On uniqueness in Candecomp/Parafac," *Psychometrika*, vol.67, no.3, pp. 399–409, Sept. 2002.
- [9] J. A. Rhodes, "A concise proof of kruskal's theorem on tensor decomposition," *Linear Algebra and its Applications*, vol.432, pp. 1818–1824, 2010.
- [10] P. Comon, "Tensors, usefulness and unexpected properties," in *15th IEEE SSP Workshop*, Cardiff, UK, Aug. 31 - Sep. 3 2009, pp. 781–788, keynote. hal-00417258.
- [11] L-H. Lim and P. Comon, "Multiarrray signal processing: Tensor decomposition meets compressed sensing," *Compte-Rendus Mec. Acad. Sci.*, vol. 338, no.6, pp. 311–320, June 2010, hal-00512271.
- [12] R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," *IEEE Trans. Information Theory*, vol.49, no.12, pp. 3320–3325, Dec. 2003.
- [13] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via l1 minimization," *Proc. Nat. Acad. Sci.*, vol.100, no.5, pp. 2197–2202, Mar. 2003.
- [14] L-H. Lim and P. Comon, "Blind multilinear identification," submitted. arxiv:1212.6663.
- [15] P. Comon, "Estimation multivariable complexe," *Traitement du Signal*, vol.3, no.2, pp. 97–101, 1986.
- [16] A. Hjørungnes and D. Gesbert, "Complex-valued matrix differentiation: Techniques and key results," *IEEE Trans. Sig. Proc.*, vol.55, no.6, pp. 2740–2746, June 2007.
- [17] P. Tichavsky, A. H. Phan and Z. Koldovsky, "Cramer-Rao Induced Bounds for Candecomp/Parafac tensor decomposition," *IEEE Trans. Sig. Proc.*, vol.61, no.8, pp. 1986–1997, 2013.